



Revue de linguistique, psycholinguistique et informatique

<http://discours.revues.org/>

## A Methodology for the Identification of Topic Transitions in Interaction

---

Marine Riou

Prismes | LLF

Université Sorbonne Nouvelle – Paris 3 | Université Paris Diderot – Paris 7

[marine.riou@univ-paris3.fr](mailto:marine.riou@univ-paris3.fr)

.....  
Marine Riou, « A Methodology for the Identification of Topic Transitions in Interaction », *Discours* [En ligne], 16 | 2015, mis en ligne le 4 septembre 2015.

.....  
URL : <http://discours.revues.org/8997>

.....  
Titre du numéro : *Varia*  
Coordination : Lydia-Mai Ho-Dac et Frédéric Landragin

**revues.org**  
CENTRE POUR L'ÉDITION ÉLECTRONIQUE OUVERTE  
CENTRE FOR OPEN ELECTRONIC PUBLISHING

 discours

 Presses  
universitaires  
de Caen



# A Methodology for the Identification of Topic Transitions in Interaction

---

Marine Riou

Prismes | LLF

Université Sorbonne Nouvelle – Paris 3 | Université Paris Diderot – Paris 7

.....

Many analysts use the notion of “topic” to describe segments of discourse, whether written or spoken, and many regret that even though a considerable amount of work has been devoted to understanding and defining the notion of discourse topic, the analyst is often left without a robust methodology to apply to their own data. This paper uses a mixed-methods approach that draws on the qualitative-oriented theoretical frameworks of Conversational Analysis and Interactional Linguistics, and combines them with quantitative methods used in other sub-fields of linguistics, such as the coding schemes and inter-rater agreement measures used in Corpus Linguistics. The goal of this study is not to provide a new understanding of “topic”, but rather to propose 1) a rich definition compiled from various earlier studies and suited to the analysis of talk-in-interaction, and 2) a systematic way to apply it to new data. The ambition of the paper is to provide a methodology for a qualitative, quantitative, or mixed-methods analysis of topic-related phenomena in interaction, as one step among many in various research protocols. It provides empirical grounds for the claim that a research methodology based on the analyst’s intuitions may be a valid and robust way to identify topic transitions – if a number of precautionary steps are taken. It also proposes a practical guide to the systematic analysis of topic in interaction.

**Keywords:** discourse topic, topic transition, talk-in-interaction, conversation analysis, coding scheme, inter-rater agreement

.....

*Un certain nombre de chercheurs font appel à la notion de « topique » pour décrire des segments de discours tant écrits qu’oraux, et beaucoup déplorent que bien qu’un nombre considérable de travaux aient porté sur la définition de cette notion, ils proposent rarement au chercheur une méthodologie robuste applicable à ses propres données. Cet article utilise une approche mixte empruntant aux cadres théoriques qualitatifs de l’analyse conversationnelle et de la linguistique interactionnelle, et les associe aux méthodes quantitatives utilisées dans d’autres branches de la linguistique, telles que les grilles de codage et les accords inter-juges mobilisés en linguistique de corpus. L’objectif de cet article n’est pas d’avancer une nouvelle conception du topique, mais plutôt de proposer 1) une définition riche du topique empruntant à plusieurs travaux existants et adaptée à l’analyse de l’interaction spontanée, et 2) une méthode pour l’appliquer de manière systématique à de nouvelles données. L’ambition est ici de proposer une méthodologie permettant tant les analyses qualitatives que quantitatives, voire mixtes, des phénomènes liés au topique en interaction, et qui constitue ainsi une étape parmi d’autres au sein de protocoles de recherche variés. Une telle base empirique permet de justifier l’argument qu’une recherche s’appuyant sur les intuitions du chercheur peut être une façon robuste et fiable d’identifier les transitions topicales, à la condition qu’un certain nombre de précautions soient prises. Cette étude propose également un guide pratique de mise en place d’une analyse systématique du topique en interaction.*

**Mots clés :** *topique discursif, transition topicale, interaction, analyse conversationnelle, grille de codage, accord inter-juges*

## 1. Introduction

- 1 The concept of discourse topic aims to capture the intuitive notion shared by many analysts that discourse is composed of several topics that are discussed in turn<sup>1</sup>. In the following example, most readers will identify l. 6 as a straightforward occurrence of a topic switch (indicated by an arrow). A list of transcription conventions can be found in Section 6, and the sound files corresponding to the examples presented can be found in the Appendix.

[1] This darn dog (SBC007, 287-303)

- |     |       |   |
|-----|-------|---|
| 1   | Alice | (..) and here another car came and rear-ended em.                                       |
| 2   | Mary  | (..) (TSK) (H) oh:[:.]  |
| 3   | Alice | [and they] ended up having to take um: (..) Peggy White (..) by helicopter to Billings. |
| 4   | Mary  | (...) man that's pretty ba:d.   |
| 5   | Alice | (.) I know.   |
| → 6 |       | (..) ↑darn (..) this darn dog keeps (..) breathing °and like° (..) dreaming             |
| 7   |       | °you know° I wonder if we should wake her up?   |
| 8   | Mary  | (.) no (.) she'll get scared and want to go outside.                                    |
| 9   |       | (..) kinda nervous you know.  |
| 10  | Alice | (..) they say you can really (..) mess up a dog by waking em up when they're dreaming.  |
| 11  | Mary  | (.) < <h> really? >   |
| 12  | Alice | (..) mhm.   |

- 2 Alice and Mary have been talking about a car crash that happened in their area, in which an acquaintance of theirs, Peggy White, was injured. Then Alice switches to a new topic l. 6 as she notices that the breathing of their dog, sleeping in the room where the recording is taking place, is getting louder. This topic transition leads them to discuss a new topic centered on sleeping dogs, and whether they can be safely woken up. Identifying this topic transition l. 6 does not constitute linguistic analysis *per se*, as it remains purely descriptive. However, this could be the first step in analyzing topic-related phenomena of talk-in-interaction, such as the structural design that participants mobilize for their topic transitions depending on their interactional project or the sequential environment. In this particular example, it might be interesting to analyze the contribution of the demonstrative *this* (l. 6)

1. I thank the anonymous reviewers and the editors for their helpful comments on an earlier draft of this article.

and how this deictic operation relates to that of turning to a new element of the extralinguistic world (the dog) and switching to this new topic at the same time.

3 This paper presents a methodology for the large-scale identification of topic transitions in talk-in-interaction. This mixed-method analysis combines a definition of topic compiled from existing work in Conversation Analysis and Interactional Linguistics, and associates it to methods used in Corpus Linguistics such as systematic coding and inter-rater agreement.

4 A lot of work has been done on the broad notion of “topic”, in a variety of theoretical frameworks. A common observation found in the literature is that this concept is widely used, but rarely defined in a way that goes beyond a lay or intuitive understanding of what it is. This led Brown and Yule (1983) to deplore that “[i]n fact, ‘topic’ should be described as the most frequently used, unexplained, term in the analysis of discourse” (Brown & Yule, 1983: 70). Goutsos (1997) concludes that there is a complete lack of consensus on the notion of topic, apart from agreement that there is no consensus: “in reviewing the work on topic and theme, one is struck by the almost total lack of consensus among linguists regarding the nature, the defining characteristics, and the scope of application of the notions employed” (Goutsos, 1997: 1). A standard criticism directed at the literature studying topic is that it often relies on a weak, vague, and intuitive definition of topic (Grobet, 2002).

5 These difficulties in the treatment of topic have led some scholars to the hypothesis that this notion should be discarded altogether as an artifact of other structuring phenomena of discourse. Thus in Relevance Theory, the notion of topic is derivative, just as coherence and cohesion are derivable from relevance (Sperber & Wilson, 1995: 217). In early work in Conversation Analysis (CA), topic was considered to be an artifact of the tying structures of interaction (Sacks, 1992). Sacks explicitly mentioned his early reluctance to consider topic as a worthwhile object of study: “I suppose I had that leeriness about ‘topic’, not by virtue of the phenomenon itself, but by virtue of that ‘topic’ would be that thing about conversation which, say, lay persons, beginning researchers, psychiatrists, etc., would feel most at home in talking about and, looking at a piece of conversation, could feel that that’s something they could start right off talking about, i.e., the ‘topics’ in it – their logic, their stupidity, the ways they were discussed, and things like that. That is to say, it would be prominently in terms of ‘topic’ that, say, ‘content analysis’ would be done” (Sacks, 1992: 752). A content analysis would indeed be contrary to the principles of the conversation analytic framework whose objective is “to uncover the tacit reasoning procedures and sociolinguistic competencies underlying the production and interpretation of talk in organized sequences of interaction” (Hutchby & Wooffitt, 1998: 14).

6 Grobet (2002) notes however that despite the different arguments to dismiss the notion of topic, an ever-growing body of work still uses this concept. She hypothesizes that “topic” could correspond to a linguistic reality, even if still poorly understood. Goutsos (1997) identified two main angles that have been taken on the notion of topic: some studies have focused on the “what” of topic (what constitutes

or not a topic), while others focus on the “how” of topic (what participants do when they “do topic”, and how they manage, perceive and cue it). Representative of the “how” approach to topic is the CA methodology, which took up the study of topic, but was careful to analyze it in conjunction with the structure of interaction, especially sequential organization (Jefferson, 1984; Button & Casey, 1984 and 1985; Maynard, 1980; Holt & Drew, 2005): “In accordance [...] with the basic CA principle of focusing on what a given bit of talk is doing rather than what it is about, [...] we will consider the various practices of speaking which conversationalists use to generate, to locate, to pursue and to resist talk on a topic. These can be thought of as practices of talk” (Sidnell, 2010: 226). Schegloff (1990: 53) warned that topic structure and sequences should remain at least partially independent.

7 I would like to demonstrate that topic can be a valuable notion when working on interactional data, not only if we want to uncover interactional structures and strategies, but also in the study of what topic is. A mixed-methods approach can allow us to work simultaneously on the “how” and the “what” of topic, on condition that it is the “how” that leads us to a “what”, i.e., only a careful analysis of how participants manage and handle topic can direct us to a working definition of what topic is in interaction.

8 The methodology presented here was developed for a larger project on the grammar and prosody of topic transition in American English conversation. One of the corpora studied was the “Santa Barbara Corpus of Spoken American English” (Du Bois et al., 2000-2005), from which all the examples presented here are extracted. Eight dyadic and casual conversations from the Santa Barbara corpus were studied, each taking place between two friends or two relatives. One fifteen-minute extract was taken from each conversation, for two hours of audio recording in total.

9 In Section 2, I give a definition of topic in talk-in-interaction compiled from existing studies and fitted to the analysis of topic transition on a large scale. Section 3 is devoted to the discussion of some issues pertaining to the analysis of topic and topic transitions in talk-in-interaction. Section 4 consists in a practical guide to identify topic transitions. It provides annotation guidelines and a discussion of inter-rater agreement as a validation procedure.

## 2. Defining discourse topic

10 In the most general way, discourse topic can be defined in terms of “aboutness” as what a portion of the interaction is about (Berthoud & Mondada, 1995; Porhiel, 2005). It is to be distinguished from the related but distinct notion of sentence topic (e.g., Lambrecht, 1994). Discourse topic is also not part of a binary opposition as is the case for the notions of topic and focus, or theme and rheme. More substantial reviews of topic in different theoretical frameworks can be found in Berthoud (1996), Goutsos (1997), Grobet (2002) and Zellers (2013). I concentrate here on putting together an understanding of discourse topic suited to interactional data and systematic analyses.

## 2.1. Topic is the center of shared attention

- 11 Topic can be defined in terms of the related notions of center of psychological or cognitive focus. Chafe (1994) defines topic as “the totality of information that is semiactive at one time” (Chafe, 1994: 128) and characterizes this information as a set of “coherently related events, states, and referents” (Chafe, 1994: 121). This definition can be linked to Gundel et al. (1993)’s work on cognitive status. With their givenness hierarchy, Gundel et al. (1993) associate different types of referents to different cognitive statuses. They link the cognitive status of being “in focus” to referents that are in short-term memory and at the current state of attention: “[t]he entities in focus at a given point in the discourse will be that partially-ordered subset of activated entities which are likely to be continued as topics of subsequent utterances” (Gundel et al., 1993: 279). Chafe (1994) draws a parallel between vision, which can be focal or peripheral, and consciousness: objects are linked to different attentional states depending on whether they are in a focal or peripheral zone of attention. Participants are aware that their co-participants have these two attention zones, and their knowledge of it influences their own production: “As they speak, they not only take account of the changing activation states of information in their own minds, but also attempt to appreciate parallel changes that are taking place in the minds of their listeners. Language is very much dependent on a speaker’s beliefs about activation states in other minds” (Chafe, 1994: 54). I retain the term “center of shared attention” to emphasize the understanding that “doing topic” is an interactional activity that has to be done jointly, as is discussed later. The following extract [2] exemplifies a very obvious shift in the cognitive focus. Pamela was telling her partner Darryl how their daughter Natalie learned about Santa Claus not being real, and then wanted to know who all the Santa Clauses she had seen in the past were:

[2] Santa Claus’s agents (SBC005, 494-516)

- |     |        |  |
|-----|--------|--|
| 1   | Pamela | I said well they’re the spirit of Santa Claus and- |
| 2   |        | (..) (H) they represent Santa Claus.               |
| 3   |        | they- (H) they’re a ↑picture [of Santa Claus.]     |
| 4   | Darryl | [they’re Santa Claus’s] agents.                    |
| 5   | Pamela | (.) @@@  |
| 6   |        | (H) they’re pictures of Santa Claus;               |
| → 7 |        | (..) < <h> is my mike on. >                        |
| 8   | Darryl | unhunh?  |
| 9   | Pamela | °oh°   |
| 10  |        | [°okay°.   |
| 11  | Darryl | [(H) it sure is.                                   |
| 12  | Pamela | (H) and.]  |
| 13  | Darryl | you just ] damn near broke the damn needle there?  |

- 14 Pamela < <h> and then she said > (..) and then she said well who  
fills the ↑stockings.
- 15 (..) and I kind of I said (..) love fills the stockings.
- 16 Darryl < <h> oh > Go:d

12 While Pamela's cognitive focus up to l. 6 is centered on the reported conversation she had with her daughter, her focal attention then switches to the microphone she is wearing for the recording. This change in the object of her attention translates to a topic transition l. 7 (*is my mike on*), which then leads Darryl to also switch his attention to the microphone. So the topic starting l. 7 is linked to a moment of shared attention. During the side sequence l. 7-13, the topic of Santa momentarily fades to a more peripheral activation state, while still remaining active. After Darryl's reassurances, Pamela goes back to the previous topic and redirects her attention to the reported conversation about Santa (l. 14), and this topic regains its status of focal center of shared attention.

## 2.2. Topic is participant- and interaction-specific

- 13 Following Brown and Yule (1983), another important component of our understanding of topic is that it is not exterior to the participants or setting: "If there is an entity identifiable as 'the topic of conversation', the analyst should consider what evidence from each individual speaker's contribution he is using to make that identification. He should also remain aware of the fact that conversation is a process and that each contribution should be treated as part of the negotiation of 'what is being talked about'. Above all, he should remember that it is speakers, and not conversations or discourses, that have 'topics'" (Brown & Yule, 1983: 94).
- 14 Mondada (2001 and 2003) also considers that doing on-topic talk is not making reference to a discourse object that is somehow autonomous or exterior to language practices. Topics are created in real time by the participants themselves. Besides the fact that it would be pointless and virtually impossible to list and verify a list of possible topics, this understanding of topic does not correspond to how participants manage it in interaction. Given the context, talking about cats then dogs might be the same topic (if discussing neighborhood issues for example) or two different topics (while visiting a pet shop for example). Only a situated analysis can help us determine what topics were discussed in a specific interaction, and more importantly, what topics were managed and oriented to as such by the participants themselves. In example [3], Richard has been explaining that he wants to keep in touch with his ex-partner Jeanie, but he hides it from his sisters because they are afraid that Jeanie is deriving too much power from the relationship:

- [3] Flowers (SBCo47, 427-447)
- 1 Richard I just wanna remain ↑friends with her=
- 2 and find out how she's doing.



- 3 Fred (..) yeah.
- 4 Richard (..) < <h> so they they tell- >
- 5 I don't tell em I call or nothing you know and uh.
- 6 (..) (TSK) and I sent her flowers last week=
- 7 < <l> I sent her flowers to work. >
- 8 Fred (TSK) < <h> y%eah[:?] >
- 9 Richard [the] d:ay before I moved out.
- 10 Fred (H: [:])
- 11 Richard [jus]t to tell her I was so:rry about everything that had happened,
- 12 and that uh (..) you know I hope (..) we could remain friends,
- 13 and that (..) God brings us back together if it was meant to be.
- 14 Fred (..) yeah.
- 15 Richard (..) and she was real happy about this.=
- 16 she said that really meant a lot to me.

15 Starting l. 6, Richard launches into a small narrative about how he sent Jeanie flowers. If we consider topic as a concept existing outside of the participants or setting, we might feel that Richard is still talking about his relationship, and that “still calling Jeanie” and “sending flowers to Jeanie” are part of one topic sequence about Richard’s relationship state. However, a situated analysis shows that Richard does not design his turn l. 6 as being a continuation of his previous turn, but rather the start of a new direction for topic development. A first series of cues suggests the end of a topic. Richard provides a summary of what has just been discussed in the current topic l. 5 (*I don’t tell them I call or nothing you know*). Interestingly enough, this topic about Richard’s sisters was brought up by Fred when he said *what what does uh (.) your (..) sisters say*. So Richard’s contribution l. 5 goes back to Fred’s topic proffer and proposes a response fitted to the original content question (Stivers, 2010) that also stands as a conclusion to the topic. Wrapping things up like this is a very common way to close off a topic (Maynard, 1980). Rather than indicating the end of a topic, other cues signal that something new is being initiated, such as the silence preceding the transition l. 6, the use of the discourse marker “and” suggesting another narrative project, and the fact that this topic transition actually develops into a full-fledged narrative. The narrative of how Richard sent flowers to Jeanie is designed as a whole with a beginning, a middle, and an end, and constitutes an interactional activity in its own right (Labov & Waletzky, 1967). Richard’s co-participant, Fred, also plays a crucial role in the constitution of this new topic about the flowers. Fred provides uptake in the form of a backchannel signal (*yeah*, l. 8). A number of prosodic cues suggest that Fred is not providing minimal uptake, but a strong invitation to say more about the matter: his *yeah* is uttered with

a laugh, high register, vocalic lengthening and a rising contour expressing appeal. All these parameters combined give an impression of amused surprise to Fred's production, and as such, sound like a ratification of the switch to the new topic.

### 2.3. Topic is co-constructed

- 16 The notion of co-construction is another key element for an interactional definition of topic (Geluykens, 1993; Mondada, 2001). In [4], Michael and his friend Jim have been discussing a voice actress who did recordings for a telephone company's automatic audio messages, indicating for example that the number called cannot be reached. Michael makes a topic transition l.2 when he suggests a new idea: a service that would be completely personalized, and not the automatic juxtaposition of pre-recorded parts:

[4] Personal computer representative (SBC017, 121-146)

- |     |         |  |
|-----|---------|--|
| 1   | Michael | (.) (H) so it's (.) (H) (.) probably the total actual (.) s- speaking that they use is pretty short. |
| → 2 |         | (..) it would be much more pleasant if they had done all the combinations though.                    |
| 3   |         | (..) °you know° call it up and there's something that actually (.) says your number.                 |
| 4   |         | (.) in total.  |
| 5   |         | (.) [you know? @@]   |
| 6   | Jim     | [yeah:=  |
| 7   |         | or] because it recognizes your phone number it automatically goes into the computer finds that.      |
| 8   | Michael | [yeah that sample.]  |
| 9   | Jim     | [and- a and names the] na:me.  |
| 10  |         | (H:) (..) < <l> thank you Mister Smith for calling Pacific Bell@ >                                   |
| 11  |         | [@@@]  |
| 12  | Michael | [yeah right.   |
| 13  |         | you kn@ow @@@ (H)]   |
| 14  | Jim     | < <l> I am your personal computer representative. >  |

- 17 This extract is a very clear example of topic co-construction, as the two participants each contribute ideas and elements in the same direction of topic development. After Michael proposes that the voice actors could have recorded all the possible combinations (l.2-4), Jim suggests that an artificial intelligence could use the phone number to address the caller by name (l.7-10) and personalize the service (l.14). Multiple consensual backchannel signals such as *yeah* (l.6, 8, 12), *right* (l.12) and

*you know* (l.5, 13) as well as shared laughter and non-competitive overlap indicate that there is alignment not only in terms of the content discussed and the stance being displayed, but also about the path of topic development. On-topic talk in this extract is a joint project to which both participants contribute.

- 18 In example [5] though, the two participants' contributions do not result in a co-constructed topic. Instead, Alice and her daughter Annette try to develop topic in two different directions, and they do not take into consideration the material contributed by each other. Annette was talking about a co-worker (*she*, l.1) who called in sick that day because of a sinus infection, and says that there is probably a virus around (l.1). Alice takes up the notion of a virus to go back to the discussion of her day at work as a pediatric nurse (l.3-4). She mentions that many patients are suffering from nasal problems, but she develops the topic in a direction that has more to do with how *she* was very busy at work and could not leave early as she was planning to.

- [5] Sinus problems (SBCo43, 161-182)
- |    |         |  |
|----|---------|--|
| 1  | Annette | she goes at least I'll get over it then.   |
| 2  | Alice   | yeah.  |
| 3  |         | (.) yeah.  |
| 4  | Annette | you know so I think it (.) just might have been something around,                          |
| 5  |         | but God that was the weirdest thing.   |
| 6  | Alice   | (.) (TSK) (H) < <h> that's what I did all day today= >                                     |
| 7  |         | I had (..) three or four different kids come up and complain of (..) nasal sinus problems? |
| 8  |         | (..) so every time one of the doc- (.) their ↑docs came on I wrote another order and.      |
| 9  | Annette | (.) I know.  |
| 10 |         | [°it's just-°]   |
| 11 | Alice   | [called Bruce] and (..) [added their name.]  |
| 12 | Annette | [it's just one of these long] going (.) flu things.  |
| 13 | Alice   | (.) Bruce wanted to go hunting today.  |
| 14 |         | < <l> and every time I call him I says > you're not getting out here early.                |

- 19 L.12, Annette makes another try at going back to the topic *she* introduced (the virus), but this is still not taken up by Alice who keeps developing a self-centered topic about the hospital. As a result, the participants in this extract do not treat topic as a joint project and they end up on parallel tracks. One sign of this is that Annette's turn l.12 (*it's just one of these long going (.) flu things*) is not really fitted to Alice's prior turn (*so every time one of the doc- (.) their docs came on I wrote another*

*order and*, l. 8), but it is actually more fitted to one of her own previous turns, which did not generate uptake from Alice (*but God that was the weirdest thing*, l. 5).

20 Since topic is co-constructed, it is very limiting to analyze a turn for itself when trying to identify topic transitions. One speaker can frame their turn as being a topic transition. However, if the new topic is not taken up in the subsequent turn(s), the proposed new topic cannot be constituted as such. Cases of aborted topics are very hard to analyze. In theory, every turn could have been interpreted as a topic transition, as every turn could have led to a subsequent development by focusing on its individual content. However, as Tannen (1984: 54) puts it: “No researcher would count every comment as a possible new topic. Rather, a topic emerges when comments are picked up and developed by the group”. Deciding that any turn is a topic transition (or not) is very difficult – unless there is clear interactional evidence that a topic is oriented to. This orientation can be positive (through topic ratification) or negative (non-ratification). In each case, participants demonstrate their awareness that a new path of topic development was suggested, and then taken up, ignored, or declined. In the absence of such evidence, it remains very problematic to consider that a specific turn *could* have been a topic transition but was not taken up.

21 An interactional analysis of topic transition is thus constrained by a double-bind. On the one hand, analyses should ideally be situated in enchrony (Enfield, 2011), i.e., conversational time, as it unfolds from the participants’ perspective, turn by turn. On the other hand, it is necessary to look at how turns subsequently develop (or drop) a potential new topic. One way to give further methodological support to analyses derived from such a double-bind is to associate them to a measure of inter-rater reliability – especially when combining qualitative analyses to systematic coding.

### 3. Analyzing topic on a large scale

22 Now that the principal components of an interactional definition of topic have been laid out (center of shared attention, participant and situation-specific, co-constructed), the question of how this definition can be used as the basis for a systematic, robust and reproducible analysis can be addressed.

23 Crow (1983) insists on the importance of shifts and frontiers between topics: “[d]efining ‘topic’ with any greater specificity than ‘what a conversation is about’ usually entails focusing on topic boundaries and shifts” (Crow, 1983: 137). In accordance with this observation, the methodology presented here focuses on the identification of occurrences when participants initiate the transition to a new topic, i.e., topic transitions.

#### 3.1. Requirements and challenges

24 A lot of work has been devoted to the notion of discourse topic, but many studies investigated only one type of topic transition or topic sequence, such as topic transitions operated over a contrastive structure (Zellers, 2013) or stepwise topic

transitions (Jefferson, 1984). Other studies took a semasiological approach and inquired into the role that a specific linguistic form can play in topic structure, such as figurative expressions (Holt & Drew, 2005), new referents (Geluykens, 1993; Gundel et al., 1993), discourse markers (Fraser, 2009), or prosody (Nakajima & Allen, 1993; Zellers, 2013).

25 Other studies were conducted on data *controlled* for topic structure. Zellers (2011) and Zellers and Post (2012) worked on read speech that had been specifically controlled for topic and phonological structure. Swerts and Geluykens' (1994) Dutch data was spontaneous and spoken, but since the interactions were task-related, the topic structure was pre-defined and corresponded to the different tasks to be performed. Geluykens' (1993) study had the merit of proposing an onomasiological approach to topic on spontaneous data, looking for different cues to topic transitions. However, it did not address the risk of circularity there is in studying markers of topic transitions (such as left-dislocations and existential *there*-constructions) when the analyst had to identify the said topic transitions in the first place – though it was noted in a later study (Swerts & Geluykens, 1994).

26 Ideally, the method used to identify topic transitions needs to be transferable to different research questions, e.g., when working on a different cue to topic structure. If we wish to base part of our analysis on the identification of topic, we not only need to identify all the topic transitions of a corpus, but we also need a way to show that this identification is reproducible. Otherwise, the highly subjective nature of the task would lead to the possibility that another researcher would have found very different results, solely based on a completely different segmentation in topic sequences.

### 3.2. Of the use of topic signals

27 Many authors have commented that the identification of topic transitions is a tricky business, as it very often relies on the analyst's intuitions. It has been suggested that transitions can be identified thanks to a number of linguistic markers. A wide array of markers of topic structure have been proposed in the literature: a concord of agreements and/or pauses (Maynard, 1980), existential structures (Berthoud, 1996; Geluykens, 1993), phatics and interjections (Berthoud & Mondada, 1995), dislocations (Grobet, 2002), clefts (Grobet, 2002), discourse markers (Horne et al., 2001) such as *so* (Bolden, 2006), new referents (Keenan Ochs & Schieffelin, 1976), etc. As far as prosodic cues are concerned, pitch parameters such as higher onsets (Nakajima & Allen, 1993), more range variation (Zellers, 2013) and fundamental frequency peaks (Swerts & Geluykens, 1994) have been identified for new topics.

28 However, this take on the identification of topic structure poses a number of problems. The first one is that very few studies have verified and quantified the actual contribution that any of these markers may play. Of course, as Mondada (2001) explains, it all comes down to what structure may be useful in one specific context, while it might not be used regularly for that action otherwise. As a result, such a

marker should not be taken as a reliable cue to topic transition. If a marker X is used in two instances, but only one is identified by the analyst as being a topic transition, then the analyst has discarded the second occurrence based solely on intuition and has not really used the marker to help in the identification process. In that case, the identification of the topic transition remained at the analyst's discretion. Consequently, this is not a good enough argument for using a time-consuming methodology that does not hold any weight against our deep-rooted intuitions about topics. If it all comes down to this, then maybe we should just embrace the intuitions we have and find a way to formalize them, by rendering them overt and reproducible.

29 The second problem is that such markers have very rarely been studied together. The fact that their individual contributions are analyzed independently of each other misses out on their interactions, while this might be a crucial component of topic signaling. A side issue is that such a list of potential markers of topic transition – until systematically verified for regularities – could remain virtually infinite, and as such would be of very little practical help to the researcher faced with the task of identifying topics in their data.

30 This take on topics also forgets that though topic marking is indeed very common, it is not obligatory, and some topic transitions are not signaled by any identifiable cue. Doing something and signaling that this action is being done are two different things. This distinction holds for topic transition: one can make a topic shift, but it does not necessarily entail that the topic shift is marked as such. Participants remain free in what they do in interaction, and depending on their interactional project, they may not wish their topic transition to be identified *as such*. Hence, looking out for topic signals may lead the analyst to miss covert topic transitions.

31 Another issue is that each of the topic signals proposed in the literature may be mobilized to do something else in interaction. For example, a higher fundamental frequency may be used to cue a new topic (Nakajima & Allen, 1993), but it can also be used during competitive overlaps (Kurtic et al., 2009). The discourse marker *anyway* can be used to return to a previous topic (Sacks, 1992), but it can also be cast as a sequence-ending device without necessarily signaling a topic shift (Park, 2010). So in the end, the task of distinguishing cases in which markers do cue topic transitions and cases in which they do not would remain at the analyst's discretion.

32 The last argument against relying on the presence of linguistic cues to identify topic transitions is that it poses a serious risk of circularity. This issue is even more problematic if the researcher is intending to analyze topic transitions in any structural way, as noted by Swerts and Geluykens (1994). Indeed, it would be very hard to justify any claims about the grammar of topic transition if the structural elements uncovered were used to identify the transitions in the first place.

33 A systematic study of topic needs a way of separating the identification of topic transitions and their linguistic analysis. Since it is virtually impossible to ignore the structural properties of the data, one solution could be to provide solid grounds that topic identification was reasonably unbiased.

34 Conversation Analysis and Interactional Linguistics provide one solution to circumvent this issue, namely, relying on participant orientation (described in Sidnell, 2013 *inter alia*). Corpus Linguistics provides a second tool with inter-rater reliability – provided it is done with a second coder who is not aware of the research question beyond the identification of topic transitions. These two methods can be used in combination to place the identification of topic transitions on more solid ground. The second one is described in the following section.

#### 4. A methodology proposal

35 This section presents the steps which were taken for the analysis of topic transitions in talk-in-interaction, which include the choice of a minimal or basic unit, as well as a practical guide to code the data and conduct an inter-agreement procedure.

##### 4.1. Choosing a minimal unit

36 One of the first tasks is to define a basic unit for one's study. Common units include the intonation unit, the utterance, the turn-at-talk and the turn-constructive unit. Once a basic unit has been chosen, for every new topic it becomes possible to analyze the design of the first unit initiating it. The choice of a basic unit of analysis also is a very important methodological decision, as it will bear on the research design, analyses, and results. One of the most easily and intuitively grasped consequences of this choice is how quantification is affected by it. To count things, one must first decide what exactly is going to be counted. More far-reaching implications arise when we try to understand a phenomenon in connection to a unit. Our conception of topic transition is affected by the fact that it could be conceived of as a phenomenon pertaining to the turn, the utterance, the intonation unit or the turn-constructive unit. What is it that *has* "topic transitions"?

37 Couper-Kuhlen (2001) identified a certain tradition personified by Chafe (1994 for example) and Du Bois (2003 for example) in which intonation units are associated to information flow and cognitive management of speech. Chafe (1994: 63) hypothesizes that "each intonation unit verbalizes the information active in the speaker's mind at its onset". According to this view, the intonation unit is the basic unit when studying discourse and speech processes. Szczepek Reed (2009) argues that the intonation unit cannot be used as a minimal unit of discourse segmentation. She recognizes that participants do produce and orient to "chunks" of speech. However, these "chunks" sometimes correspond to conventional units such as intonation units, and sometimes they do not. The production and recognition of chunks involves prosody, but also different modalities. Consequently, using the intonation unit as the basic unit of interaction is limiting. By contrast, choosing a basic unit that is interactional in nature makes it possible to avoid relying on a pre-established theoretical conception of what an intonation unit is in the first place (Szczepek Reed, 2009: 359). Zellers (2011) also argues that there is little reason to think that topic structure is inscribed in the phonology: "A few prosodic studies have investigated

the internal structure of these units [intonational units]. It seems relatively clear that this [topical] structure is not a part of the phonology of a language per se; that is, we would not expect to find topic-structure variation encoded as part of an intonational grammar. Instead, it is part of the discourse structure” (Zellers, 2011: 81-82).

- 38 Transitioning to a new topic is a conversational move, and as such can be done over the course of one *interactional* unit. The minimal unit chosen for this project is the turn-constructive unit (TCU). The TCU corresponds to a potentially complete turn-at-talk and is projected by a wide array of syntactic, semantic, pragmatic and prosodic cues (Ford et al., 1996; Selting, 2000). In the following extract, Fred is asking about Richard’s new job. The transcription layout corresponds to the transcription conventions of the Santa Barbara corpus, in which one line corresponds to one intonation unit:

[6] Training (SBCo47, 652-659)

- |   |   |         |   |
|---|---|---------|---|
|   | 1 | Richard | you know?   |
|   | 2 | Fred    | (.) yeah.   |
| → | 3 |         | (..) (H:) < <h> so homes, >                               |
|   | 4 |         | (.) uh,   |
|   | 5 |         | (..) they-  |
|   | 6 |         | (.) they put you through < <h> training all these days? > |
|   | 7 |         | [is that it?]   |
|   | 8 | Richard | [yeah.]   |

- 39 Fred initiates a new topic l. 3, but he needs four intonation units to do it. These four intonation units (l. 3-6) correspond to one TCU, as Fred’s turn could not be complete before the end of l. 6: *so homes* (l. 3), *so homes uh* (l. 3-4) or *so homes uh they* (l. 3-5) could not constitute a complete turn in this environment, as the continuative prosody transcribed with commas indicate.

- 40 From this point on, each line corresponds to a TCU in the examples provided. The segmentation in TCUs was done following the methodology described in Selting (2000). The TCU presents the practical advantage of being widely used in CA and Interactional Linguistics, and the theoretical advantage of presenting neurological (Bögels et al., 2014) and multimodal correlates (Holler & Kendrick, 2014).

- 41 Since TCUs correspond to interactional moves, they can correspond to topic transitions. In that case, the analyst will look for the TCUs initiating talk on a topic. As long as the unit chosen is a discursive and/or interactional one, another minimal unit could have been chosen, such as the Basic Discourse Unit (Degand & Simon 2009; Simon & Degand, 2011). This choice will depend on a variety of parameters such as the research question, the type of data, and the theoretical framework.



#### 4.2. A practical guide for systematic coding

42 If the analyst is working on a corpus that is not already transcribed, the data should be transcribed following established transcription conventions, such as for example the system designed by Gail Jefferson for CA (Jefferson, 2004), the Discourse Transcription system (DT or its later version DT2) established by Du Bois et al. (1992) for the Santa Barbara corpus, or the CHAT format designed for the CLAN program (MacWhinney, 2000). For more on the theoretical and practical implications inherent to any transcription, see Ochs (1979), Du Bois (1991), and Edwards and Lampert (1993) inter alia. It is strongly recommended to transcribe one's corpus using a program that can align the transcription to the audio or video file, such as CLAN (MacWhinney, 2000). A time-aligned transcription allows the analyst to go back to the sound file constantly and to avoid working on the transcript alone.

43 The next recommended step is to export the whole corpus into a spreadsheet where all the coding can be done. To facilitate later treatment, a very effective setup is to copy and paste the whole corpus in only one spreadsheet, even if the corpus is made up of a collection of different interactions. All the other columns can be devoted to the coding of the data. It is advisable to reserve a number of columns for metadata such as the name or reference number of each conversation as well as speaker identification. In that respect, Gries (2013: 20-26) gives very explicit recommendations. The following table is a simplified version of the coding scheme used for the present study:

	A	B	C	D	E	F
	TCU	Conversation	Participant	Topic transition	Type of topic transition	Ratification
1	hi sweetie	SBCo34	Karen	o	NA	NA
2	(..) hey	SBCo34	Scott	o	NA	NA
3	(.) (THROAT)	SBCo34	Scott	o	NA	NA
4	sweetie fumpions	SBCo34	Karen	o	NA	NA
5	(..) this is kinda open	SBCo34	Karen	i	disjunctive	i
6	(..) yep	SBCo34	Scott	o	NA	NA
7	(..) how was work	SBCo34	Scott	i	disjunctive	i

Table 1. Coding spreadsheet for the analysis of topic transition

44 In that type of setup, one conversation can be read in its entirety from top to bottom in the leftmost column, and each row corresponds to one minimal unit.

Table 1 shows the beginning of the conversation SBCo34 and starts with an opening sequence consisting of a greeting-greeting sequence (1.1-2). The coding scheme exemplified here implies that each TCU of the corpus needs to be coded for the same coding categories. Column D is the focus of this paper, as it concerns the identification of topic transitions. Each TCU was coded 1 if it corresponded to the first sign of a switch to a new topic (such as the TCUs in rows 5 and 7), or coded 0 if it did not. Two other coding categories (columns E and F) were related to topic management but remain outside the scope of this paper. They are only shown here for illustration purposes: the type of topic transition, which could be indicated as being a disjunctive or a stepwise transition (Jefferson, 1984; Holt & Drew, 2005); and topic ratification (i.e., whether the topic transition was validated by the co-participant's orientation and contribution).

45 Coding one's interactional data systematically makes it possible to use analytical tools developed in Corpus Linguistics. More specifically, the spreadsheet setup presented in Table 2 is geared towards multifactorial usage-feature analysis (Glynn, 2014). In this perspective, usage features are uncovered through the systematic manual coding of large bodies of data. Each usage feature can then be treated as a variable and multivariate statistics can be conducted.

46 It cannot be stressed strongly enough that any systematic coding should be accompanied by a careful qualitative and situated analysis of the data. Ideally, the interplay between qualitative and quantitative analyses should remain constant throughout the research process. Indeed, only a careful qualitative analysis can provide the coding categories adequate to a specific research question and data, and only a qualitative mindset can give meaning to the results obtained through quantification. On the other hand, only an even minimal level of quantification can give rise to generalizations that go beyond single case analyses. CA has always been wary of quantification (Schegloff, 1993), as the close analysis of the uniqueness of specific cases is at the heart of the CA methodology. However, recent studies have proven that CA can benefit from quantitative methods without endangering its core beliefs (Stivers et al., 2009; Stivers & Enfield, 2010). Stivers (2015: 16) makes a strong case that "formal coding can provide a second story of the analytic house, thus improving the view and reach of CA research". Hence, a mixed-methods approach to interaction can provide researchers with a more comprehensive and varied set of analytical tools.

#### 4.3. Topic transition identification and inter-rater agreement

47 An inter-rater reliability procedure measures the agreement between several raters or coders. It can help evaluate the reliability and relative objectivity of a coding scheme. For this study, one category of a more comprehensive coding scheme was submitted to this procedure: the identification of topic transitions. The choice that the coders had was binary: they had to decide for each TCU whether it corresponded to the initiation of a new topic or not.

The standard procedure for an inter-rater reliability check in linguistics is to carry it out on at least 10% of the corpus. Two 15-minute conversations were subjected to this procedure, which corresponds to 25% of the 2-hour corpus used in the study.

The two coders were the author and a colleague with a comparable training in the linguistic analysis of English data and non-native high level of proficiency in English (C2 level in the CEFR [Common European Framework of Reference for Languages]). After giving the second coder written instructions on how to code the data and a list of the transcription conventions used in the Santa Barbara corpus, we coded together and at the same time a portion of the corpus which was to serve as a demonstration, and which was not included in the 25% of the corpus tested for inter-rater reliability. After the demonstration and training phase, we coded separately the two 15-minute conversations selected for inter-rater agreement, and then we compared our coding.

Table 2 presents a model for the annotation spreadsheet which can be given to the second coder, along with the corresponding sound files and/or time-aligned transcripts. Columns A, B and C contain the content of the conversation(s) as well as metadata, and the identification of topic transitions can be done in column D. Hence the second coder has no indication as to which TCUs the first coder identified as topic transitions.

	A	B	C	D
	TCU	Conversation	Participant	Topic transition Coder #2
1	hi sweetie	SBCo34	Karen	
2	(..) hey	SBCo34	Scott	
3	(.) (THROAT)	SBCo34	Scott	
4	sweetie fumptions	SBCo34	Karen	
5	(..) this is kinda open	SBCo34	Karen	
6	(..) yep	SBCo34	Scott	
7	(..) how was work	SBCo34	Scott	

Table 2. Coding spreadsheet for the second coder

The coding spreadsheet (Table 2) was accompanied with coding instructions detailing the content of each column and the numerical code to be used for topic transition identification (Table 3).

After the two coders have annotated the data separately, their ratings can be combined in the same file. For illustration purposes, Table 4 presents the fictitious results of an inter-rater agreement conducted on the short extract presented earlier in Tables 1 and 2.

A	TCU	The conversation is segmented in turn-constructional units (TCUs), an interactional unit. This column contains all the TCUs of the conversation, in the order in which they are produced.
B	Conversation	The name of the conversation in which this TCU was produced.
C	Participant	The name of the participant who produces this TCU.
D	Topic transition	Whether there is a topic transition or not in this TCU: o = no topic transition, i.e. this TCU is about the same topic as the previous TCU; 1 = topic transition, i.e. this TCU is about a new topic, different from the topic under discussion in the previous TCU.

Table 3. Coding guidelines for the second coder

	A	B	C	D	E	F
	TCU	Conversation	Participant	Topic transition Coder #1	Topic transition Coder #2	Agreements
1	hi sweetie	SBCo34	Karen	o	1	o
2	(..) hey	SBCo34	Scott	o	o	1
3	(.) (THROAT)	SBCo34	Scott	o	o	1
4	sweetie fumptions	SBCo34	Karen	o	o	1
5	(..) this is kinda open	SBCo34	Karen	1	1	1
6	(..) yep	SBCo34	Scott	o	o	1
7	(..) how was work	SBCo34	Scott	1	1	1

Table 4. Agreements between the two coders

53 Column F specifies whether the two coders agreed or disagreed on the status of each TCU as a topic transition or continuity. The code 1 indicates that they agreed, whether they both identified a TCU as a topic transition (1.5), or when they both identified a TCU as *not* being a transition (1.2). The code o indicates that they disagreed, i.e., whenever they did not use the same code, as for example 1.1 where coder #1 did not see a topic transition but coder #2 did. Based on the results of column F, it is then possible to calculate the amount of agreements and disagreements on topic transitions.

54 An agreement coefficient compares the probability of the observed agreement between coders ( $\text{Pr}(a)$ ) to the probability of a random agreement ( $\text{Pr}(e)$ ), and it is based on the following formula:

$$\kappa = \frac{\text{Pr}(a) - \text{Pr}(e)}{1 - \text{Pr}(e)}$$

55 For an inter-rater reliability check with two coders, the statistical test used is Cohen's *kappa* ( $\kappa$ ). In the case of a binary coding scheme (i.e., the coders could only choose between two options), a random agreement would correspond to a  $\text{Pr}(e)$  of 0.5. A *kappa* of 0 would correspond to a chance agreement, and 1 to a perfect agreement. When interpreting the result of an inter-rater reliability check, it has become a convention to refer to the following scale (Landis & Koch, 1977):

<i>Kappa</i>	Agreement
< 0	Less than chance agreement
0.01–0.20	Slight agreement
0.21–0.40	Fair agreement
0.41–0.60	Moderate agreement
0.61–0.80	Substantial agreement
0.81–0.99	Almost perfect agreement

56 However, such a scale is purely indicative and depends greatly on the field and phenomenon analyzed. In practice, the standard threshold in most fields, including computational linguistics, is 0.8. Artstein and Poesio (2008) recommend lowering the acceptable threshold to 0.7 in the case of discourse studies.

57 Out of the 1,130 TCUs coded for this procedure, 1,092 were coded identically by the two coders, which corresponds to an agreement rate of 97% (95% Confidence Interval [CI]: 95–97%). Cohen's *kappa* as computed in the statistical software R (2014) with the function `kappa2()` (`irr` package) reads  $\kappa = 0.73$ . This agreement can be considered high enough, as it exceeds the minimum threshold of 0.7 and shows the two coders did not differ too much from a perfect agreement.

58 The two coders agreed on 57 Transitions and 1,035 Continuities, for a total of 1,092 agreements (see Table 5). Disagreements amounted to 38. The second coder identified 16 Transitions that were not identified by the author ("false negatives"), and the author identified 22 Transitions that the second coder coded as Continuities ("false positives"). So the author was more liberal and the second coder more conservative.

		Coder 1		
		Transition	Continuity	Total
Coder 2	Transition	57	16	73
	Continuity	22	1,035	1,057
	Total	79	1,051	1,130

Table 5. Confusion matrix for the inter-rater agreement

59 Disagreements corresponded to the identification of transitions but not to their exact location. There were only a trivial number of disagreements about the exact location of cut-off points between topics. Whenever the two coders agreed on the presence of a transition, they also agreed on which TCU initiated this transition.

60 Most disagreements were connected to one type of environment. In the environment of a narrative (or “larger project”, Selting, 2000), the second coder tended not to perceive as a topic transition a TCU introducing a different part of the narrative, while the first coder did. In [7], Alice has been describing the first day at work of her new manager. It has not been an altogether pleasant day for the newcomer, as she had to deal with a series of inconveniences which Alice describes in turn. Each inconvenience is treated as one element in a long narrative explaining how “it has not been a good day” (SBCo43, 92). After a sequence about the manager’s pants which did not fit (l.1-4), Alice launches into a new part of this narrative about an attaché case which could not be opened (l.5). The TCU l.5 is a typical example for which the most liberal coder identified a topic transition and the most conservative coder did not.

[7] A brand new attaché case (SBCo43, 36-50)

- |     |         |   |
|-----|---------|---|
| 1   | Alice   | (.) she goes < <h> and I didn’t bother > to try em on before<br>(.) I left.                   |
| 2   |         | (H) so she’s < <l> all of em (.) are not short enough. >                                      |
| 3   |         | she’s (.) these are the shorter one and they’re about<br>two inches too@ loʃ:ng@.             |
| 4   | Annette | [@ (H) oh:@ ma:n.]  |
| → 5 | Alice   | (.) (H) < <h> and then] > (.) she brought- bought a brand<br>new attaché case.                |
| 6   |         | (..) and yesterday was the first day she ↑used it.  |
| 7   |         | (H) put a bunch of stuff in it to read (H) went home last<br>night and couldn’t get it open=. |
| 8   |         | the lock would not open.  |
| 9   |         | (..) < <h> and she’s > (H) I don’t think it came with a key<br>and it’s this- (.) k- a key.   |

61 It is indeed open to debate whether within a narrative, each part corresponds to a new topic or sub-topic, or whether only the whole narrative corresponds to a single topic. This issue might not have arisen if the coding had not been binary. As the coders had to indicate whether each TCU was a Transition or Continuity, this did not leave any latitude to differentiate between different topics and different sub-topics connected to a macro-topic. A finer-grained coding scheme could address this issue more adequately.

62 Such a result is also interesting in that it can provide the analyst with insights as to which transitions are more consensual. Transitions on which the two coders agreed may exhibit characteristics that make them more prototypical transitions. By contrast, non-consensual transitions may also be interesting in that they might be less linguistically marked, more covert, or less framed as being transitions by the participants – depending on the angle from which we look at the data.

63 Another caveat concerning the inter-rater agreement has to do with the fact that even if  $\kappa$  is a chance-corrected measure of agreement (contrary to a simple percentage of agreement), it does not prevent issues of prevalence: “If a disproportionate amount of the data falls under one category, then the expected agreement is very high, so in order to demonstrate high reliability an even higher observed agreement is needed” (Artstein & Poesio, 2008: 573). In the case of the present study, the number of topic transitions is so low in the data (compared to continuing TCUs) that the two coders were bound to agree that most TCUs are not topic transitions – which skews the results in favor of the coding scheme. In this case, one solution can be to look at the less common category, which here is topic transition (vs. topic continuity). In the two conversations subjected to this test, the author found 79 topic transitions and the second coder found 73. This does not mean that the two coders found the exact same ones but for 6 (79-73). However, among the 79 TCUs identified as topic transitions by the author, 57 of them were coded similarly by the second coder, which corresponds to a 72% (95% CI: 61-81%) agreement rate. It would not be acceptable to compute a chance-corrected measure here, as a *kappa*-style coefficient assumes that the raters coded the data independently. In the present scenario, we would not be assessing how well the two coders agreed with each other, but how well the second coder agreed with respect to the author’s coding. However, this agreement rate on the author’s own identification of topic transitions, together with the chance-corrected  $\kappa$  of 0.73 for the whole procedure, suggest that topic identification is reasonably reproducible.

64 By precisely indicating to what extent the two coders agreed on the identification of topic transitions, an inter-rater agreement formalizes practices that are not always brought forward by researchers, and it allows results to be tested for replication. Besides, it is always possible to refine the coding scheme and directions as many times as it takes for the two coders to reach a better agreement, on the condition that a new set of data is used each time.

65 Even though topic transition identification has always been deemed highly subjective and thus to be avoided, it appears that though it certainly is, this does not entail that it cannot be formalized and used as part of a research method. Provided that researchers explicitly state how they came by their segmentation and provide a measure such as inter-rater agreement, identifying topics in talk-in-interaction can be part of a research methodology.

## 5. Conclusion

66 An important caveat that needs to be stressed again is that the proposed methodology is only a research tool, and it should not affect our conception of topic in talk-in-interaction. The methodology presented here makes simplifications that cannot hold when analyzing data in context. Contrary to what the coding methodology

implies, topic is not linear, and nothing shows that one topic transition would correspond to exactly one TCU. However, this simplification proves to be a good enough place to get started, especially when handling large amounts of data with a view to conducting even basic quantitative analyses. This segmentation in topic sequences can fruitfully parallel and intersect with other planes of analysis, such as the close sequential analysis required in CA. Whether the researcher wishes to study topic itself or another phenomenon of talk-in-interaction, s/he cannot maintain this view of topic when performing the necessary qualitative analysis of the data. This paper has argued that if the necessary precautions are taken, it can be a sound and promising place to start nonetheless, especially if it allows researchers to ask topic-related questions on a large scale.

## 6. Transcription conventions

67 The transcription conventions used in this paper correspond to the DT2 system set up by Du Bois et al. (1992) for the Santa Barbara corpus. However, since the data from the Santa Barbara corpus were segmented in TCUs for this study, some intonation contours were suppressed when several intonation units of the original transcripts were combined in one TCU. Additional prosodic notations used by Szczepek Reed (2011) were added. The following list explains the symbols used in this paper:

(.)	very short pause
(..)	short/medium pause
(...)	longer pause
=	latching
:	lengthening
[ ]	overlap with following turn
[ ]	overlap with previous turn
(H)	inhalation
(TSK)	click
@	laugh pulse
wo@rd	laughing word
(%)	glottal stop, creak
w%oord	glottalized word
°word°	piano, attenuated speech
↑	pitch step-up
< <l> >	low pitch register
< <h> >	high pitch register
.	terminative contour
,	continuative contour
?	appeal contour
-	truncated contour



## References

- ARTSTEIN, R. & POESIO, M. 2008. Inter-Coder Agreement for Computational Linguistics. *Computational Linguistics* 34 (4): 555-596. Available online: <http://www.aclweb.org/anthology/J08-4004>.
- BERTHOUD, A.-C. 1996. *Paroles à propos: approche énonciative et interactive du topic*. Paris – Gap: Ophrys.
- BERTHOUD, A.-C. & MONDADA, L. 1995. Traitement du topic, processus énonciatifs et séquences conversationnelles. *Cahiers de linguistique française* 17: 205-228. Available online: <http://clf.unige.ch/display.php?numero=17&idFichier=223>.
- BÖGELS, S., MAGYARI, L. & LEVINSON, S. 2014. EEG Correlates of Processes Related to Turn-Taking in an Interactive Quiz Paradigm. In *4th International Conference on Conversation Analysis – ICCA-14* (25-29 June 2014, University of California at Los Angeles).
- BOLDEN, G.B. 2006. Little Words That Matter: Discourse Markers “So” and “Oh” and the Doing of Other-Attentiveness in Social Interaction. *Journal of Communication* 56 (4): 661-688.
- BROWN, G. & YULE, G. 1983. *Discourse Analysis*. Cambridge – London: Cambridge University Press.
- BUTTON, G. & CASEY, N. 1984. Generating Topic: The Use of Topic Initial Elicitors. In J.M. ATKINSON & J. HERITAGE (eds.), *Structures of Social Action: Studies in Conversation Analysis*. Cambridge – Paris: Cambridge University Press – Éditions de la Maison des sciences de l’homme: 167-190.
- BUTTON, G. & CASEY, N. 1985. Topic Nomination and Topic Pursuit. *Human Studies* 8 (1): 3-55.
- CHAFE, W. 1994. *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago – London: University of Chicago Press.
- COUPER-KUHLEN, E. 2001. Intonation and Discourse: Current Views from Within. In D. SCHIFFRIN, D. TANNEN & H.E. HAMILTON (eds.), *The Handbook of Discourse Analysis*. Malden – Oxford – Carlton: Blackwell: 13-34.
- CROW, B.K. 1983. Topic Shifts in Couples’ Conversations. In R.T. CRAIG & K. TRACY (eds.), *Conversational Coherence: Form, Structure, and Strategy*. London: Sage: 136-156.
- DEGAND, L. & SIMON, A.C. 2009. Mapping Prosody and Syntax as Discourse Strategies: How Basic Discourse Units Vary across Genres. In D. BARTH-WEINGARTEN, N. DEHÉ & A. WICHMANN (eds.), *Where Prosody Meets Pragmatics*. Bingley: Emerald: 79-105.
- DU BOIS, J.W. 1991. Transcription Design Principles for Spoken Discourse Research. *Pragmatics* 1 (1): 71-106.
- DU BOIS, J.W. 2003. Discourse and Grammar. In M. TOMASELLO (ed.), *The New Psychology of Language: Cognitive and Functional Approaches to Language Structure*. Mahwah – London: L. Erlbaum. Vol. 2: 47-88.
- DU BOIS, J.W. et al. (eds.) 1992. *Santa Barbara Papers in Linguistics: Discourse Transcription* 4. Santa Barbara: Department of Linguistics at the University of California.

- DU BOIS, J.W. et al. 2000-2005. *Santa Barbara Corpus of Spoken American English, Parts 1-4*. Philadelphia: Linguistic Data Consortium. DVD.
- EDWARDS, J.A. & LAMPERT, M.D. (eds.) 1993. *Talking Data: Transcription and Coding in Discourse Research*. Hillsdale – London: L. Erlbaum.
- ENFIELD, N.J. 2011. Sources of Asymmetry in Human Interaction: Enchrony, Status, Knowledge and Agency. In T. STIVERS, L. MONDADA & J. STEENSIG (eds.), *The Morality of Knowledge in Conversation*. Cambridge – New York: Cambridge University Press: 285-312.
- FORD, C.E., FOX, B.A. & THOMPSON, S.A. 1996. Practices in the Construction of Turns: The “TCU” Revisited. *Pragmatics* 6 (3): 427-454.
- FRASER, B. 2009. Topic Orientation Markers. *Journal of Pragmatics* 41 (5): 892-898.
- GELUYKENS, R. 1993. Topic Introduction in English Conversation. *Transactions of the Philological Society* 91 (2): 181-214.
- GLYNN, D. 2014. Techniques and Tools: Corpus Methods and Statistics for Semantics. In D. GLYNN & J.A. ROBINSON (eds.), *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy*. Amsterdam – Philadelphia: J. Benjamins: 307-341.
- GOUTSOS, D. 1997. *Modeling Discourse Topic: Sequential Relations and Strategies in Expository Text*. Norwood: Ablex.
- GRIES, S.T. 2013. *Statistics for Linguistics with R: A Practical Introduction*. Berlin – New York: Mouton de Gruyter.
- GROBET, A. 2002. *L'identification des topiques dans les dialogues*. Brussels – Paris: Duculot.
- GUNDEL, J.K., HEDBERG, N. & ZACHARSKI, R. 1993. Cognitive Status and the Form of Referring Expressions in Discourse. *Language* 69 (2): 274-307.
- HOLLER, J. & KENDRICK, K. 2014. Gaze and the Observation of Turn-Taking in Triadic Face-to-Face Interaction. In *4th International Conference on Conversation Analysis – ICCA-14* (25-29 June 2014, University of California at Los Angeles).
- HOLT, E. & DREW, P. 2005. Figurative Pivots: The Use of Figurative Expressions in Pivotal Topic Transitions. *Research on Language and Social Interaction* 38 (1): 35-61.
- HORNE, M. et al. 2001. Cue Words and the Topic Structure of Spoken Discourse: The Case of Swedish Men “But”. *Journal of Pragmatics* 33 (7): 1061-1081.
- HUTCHBY, I. & WOOFFITT, R. 1998. *Conversation Analysis: Principles, Practices, and Applications*. Cambridge – Malden: Polity Press.
- JEFFERSON, G. 1984. On Stepwise Transition from Talk about a Trouble to Inappropriately Next-Positioned Matters. In J.M. ATKINSON & J. HERITAGE (eds.), *Structures of Social Action: Studies in Conversation Analysis*. Cambridge – Paris: Cambridge University Press – Éditions de la Maison des sciences de l'homme: 191-222.
- JEFFERSON, G. 2004. Glossary of Transcript Symbols with an Introduction. In G.H. LERNER (ed.), *Conversation Analysis: Studies from the First Generation*. Amsterdam – Philadelphia: J. Benjamins: 13-31.
- KEENAN OCHS, E. & SCHIEFFELIN, B.B. 1976. Topic as a Discourse Notion: A Study of Topic in the Conversations of Children and Adults. In C.N. LI (ed.), *Subject and Topic*. New York – San Francisco – London: Academic Press: 335-384.

- KURTIC, E., BROWN, G.J. & WELLS, B. 2009. Fundamental Frequency Height as a Resource for the Management of Overlap in Talk-in-Interaction. In D. BARTH-WEINGARTEN, N. DEHÉ & A. WICHMANN (eds.), *Where Prosody Meets Pragmatics*. Bingley: Emerald: 183-203.
- LABOV, W. & WALETZKY, J. 1967. Narrative Analysis. In J. HELM (ed.), *Essays on the Verbal and Visual Arts*. Seattle – London: University of Washington Press: 74-104 [reprinted in 1997 in *Journal of Narrative and Life History* 7: 3-38].
- LAMBRECHT, K. 1994. *Information Structure and Sentence Form: Topic, Focus and the Mental Representations of Discourse Referents*. Cambridge: Cambridge University Press.
- LANDIS, J.R. & KOCH, G.G. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33 (1): 159-174.
- MACWHINNEY, B. 2000. *The CHILDES Project: Tools for Analyzing Talk*. Mahwah – London: L. Erlbaum [3rd edition].
- MAYNARD, D.W. 1980. Placement of Topic Changes in Conversation. *Semiotica* 30 (3-4): 263-290.
- MONDADA, L. 2001. Gestion du topic et organisation de la conversation. *Cadernos de estudos lingüísticos* 41: 7-35. Available online: <http://revistas.iel.unicamp.br/index.php/cel/article/viewFile/1771/1343>.
- MONDADA, L. 2003. Parler topical et organisation séquentielle: l'apport de l'analyse conversationnelle. *Verbum* 25 (2): 193-219.
- NAKAJIMA, S. & ALLEN, J.F. 1993. A Study on Prosody and Discourse Structure in Cooperative Dialogues. *Phonetica* 50 (3): 197-210.
- OCHS, E. 1979. Transcription as Theory. In E. OCHS & B.B. SCHIEFFELIN (eds.), *Developmental Pragmatics*. New York – London: Academic Press: 43-72.
- PARK, I. 2010. Marking an Impasse: The Use of *anyway* as a Sequence-Closing Device. *Journal of Pragmatics* 42 (12): 3283-3299.
- PORHIEL, S. 2005. Les marqueurs de thématisation: des thèmes phrastiques et textuels. *Travaux de linguistique* 51: 55-84.
- R DEVELOPMENT CORE TEAM 2014. *The R Project for Statistical Computing*. Version 3.0.2 (2013-09-25). Available online: <http://www.R-project.org>.
- SACKS, H. 1992. *Lectures on Conversation*. JEFFERSON, G. (ed.). Oxford – Cambridge: Blackwell.
- SCHEGLOFF, E.A. 1990. On the Organization of Sequences as a Source of "Coherence" in Talk-in-Interaction. In B. DORVAL (ed.), *Conversational Organization and Its Development*. Norwood: Ablex: 51-77.
- SCHEGLOFF, E.A. 1993. Reflections on Quantification in the Study of Conversation. *Research on Language and Social Interaction* 26 (1): 99-128.
- SELTING, M. 2000. The Construction of Units in Conversational Talk. *Language in Society* 29 (4): 477-517.
- SIDNELL, J. 2010. *Conversation Analysis: An Introduction*. Chichester: Wiley-Blackwell.
- SIDNELL, J. 2013. Basic Conversation Analytic Methods. In J. SIDNELL & T. STIVERS (eds.), *The Handbook of Conversation Analysis*. Oxford: Wiley-Blackwell: 77-99.

- SIMON, A.-C. & DEGAND, L. 2011. L'analyse en *unités discursives de base*: pourquoi et comment? *Langue française* 170: 45-59.
- SPERBER, D. & WILSON, D. 1995. *Relevance: Communication and Cognition*. Oxford – Malden: Blackwell.
- STIVERS, T. 2010. An Overview of the Question-Response System in American English Conversation. *Journal of Pragmatics* 42 (10): 2772-2781.
- STIVERS, T. 2015. Coding Social Interaction: A Heretical Approach in Conversation Analysis? *Research on Language and Social Interaction* 48 (1): 1-19.
- STIVERS, T. et al. 2009. Universals and Cultural Variation in Turn-Taking in Conversation. *Proceedings of the National Academy of Sciences* 106 (26): 10587-10592.
- STIVERS, T. & ENFIELD, N.J. 2010. A Coding Scheme for Question-Response Sequences in Conversation. *Journal of Pragmatics* 42 (10): 2620-2626.
- SWERTS, M. & GELUYKENS, R. 1994. Prosody as a Marker of Information Flow in Spoken Discourse. *Language and Speech* 37 (1): 21-43.
- SZCZEPEK REED, B. 2009. Units of Interaction: “Intonation Phrases” or “Turn Constructional Phrases”? In H.-Y. YOO & E. DELAIS-ROUSSARIE (eds.), *Actes d'IDP 2009 – Interface, discours et prosodie*. 351-363. Available online: [http://makino.linguist.jussieu.fr/idp09/docs/IDP\\_actes/Articles/szczepek-reed.pdf](http://makino.linguist.jussieu.fr/idp09/docs/IDP_actes/Articles/szczepek-reed.pdf).
- SZCZEPEK REED, B. 2011. *Analysing Conversation: An Introduction to Prosody*. Basingstoke – New York: Palgrave Macmillan.
- TANNEN, D. 1984. *Conversational Style: Analyzing Talk among Friends*. Norwood: Ablex.
- ZELLERS, M. 2011. *Prosodic Detail and Topic Structure in Discourse*. PhD thesis. University of Cambridge.
- ZELLERS, M. 2013. Prosodic Variation for Topic Shift and Other Functions in Local Contrasts in Conversation. *Phonetica* 69 (4): 231-253.
- ZELLERS, M. & POST, B. 2012. Combining Formal and Functional Approaches to Topic Structure. *Language and Speech* 55 (1): 119-139.

## Appendix

### Sound files corresponding to the examples

- [https://archive.org/download/o2Riou7SoundFiles/o2Riou\\_Example1\\_dog.wav](https://archive.org/download/o2Riou7SoundFiles/o2Riou_Example1_dog.wav)
- [https://archive.org/download/o2Riou7SoundFiles/o2Riou\\_Example2\\_santa.wav](https://archive.org/download/o2Riou7SoundFiles/o2Riou_Example2_santa.wav)
- [https://archive.org/download/o2Riou7SoundFiles/o2Riou\\_Example3\\_flowers.wav](https://archive.org/download/o2Riou7SoundFiles/o2Riou_Example3_flowers.wav)
- [https://archive.org/download/o2Riou7SoundFiles/o2Riou\\_Example4\\_computer.wav](https://archive.org/download/o2Riou7SoundFiles/o2Riou_Example4_computer.wav)
- [https://archive.org/download/o2Riou7SoundFiles/o2Riou\\_Example5\\_sinus.wav](https://archive.org/download/o2Riou7SoundFiles/o2Riou_Example5_sinus.wav)
- [https://archive.org/download/o2Riou7SoundFiles/o2Riou\\_Example6\\_training.wav](https://archive.org/download/o2Riou7SoundFiles/o2Riou_Example6_training.wav)
- [https://archive.org/download/o2Riou7SoundFiles/o2Riou\\_Example7\\_case.wav](https://archive.org/download/o2Riou7SoundFiles/o2Riou_Example7_case.wav)

These 7 sound files are available online: <https://archive.org/details/o2Riou7SoundFiles>.